

Washington State Archives, Digital Archives

The evolution of ingestion

Presented by:

Daniel Waterbly

dan.waterbly@sos.wa.gov

September 2010, Cheney WA

Archive This!

- Supports many data formats
- Validates raw data on partner's machine
- Structures incoming data into common file structure
- Transfer data to hard drive or allows data to be uploaded over FTPS
- Fingerprints all incoming data before we receive it at the Digital Archives
- Gateway to Ingestion System

Archive This

Archive This! - Washington State Archives, Digital Archives

Archive This!

User Name

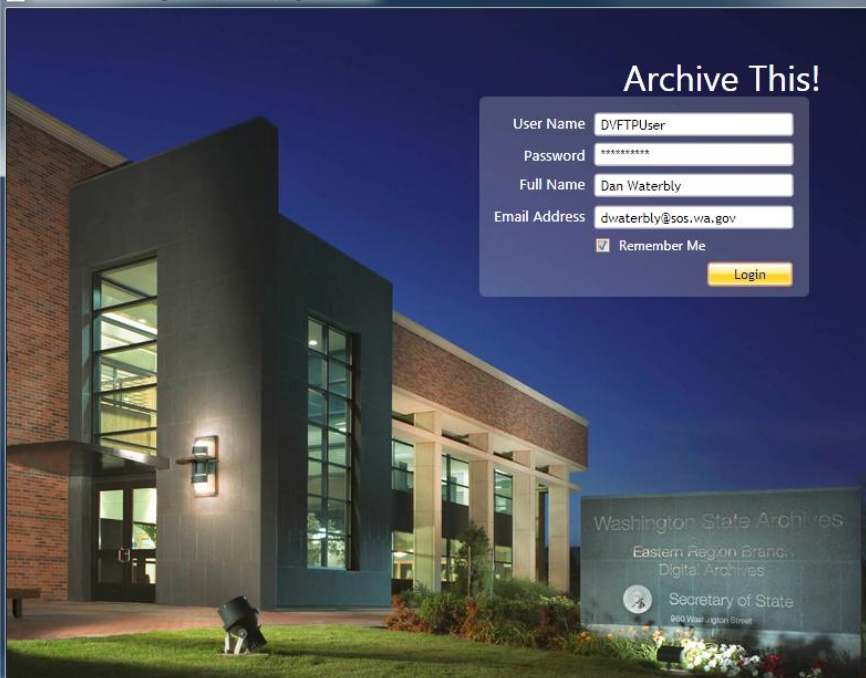
Password

Full Name

Email Address

☒ Remember Me

Login



Archive This! - Washington State Archives, Digital Archives

Select a Data Format

Please select from the list below the type of recording system data you will be sending to the Digital Archives. When you are finished click the continue button.

Data Format

Are your records from the holdings of the Washington State Archives? ☐ Yes ☒ No

Next

Archive This

Archive This! - Washington State Archives, Digital Archives

What Are You Sending Us?

Please use the form below to select the records you will be sending to the Digital Archives, if the appropriate records series is not displayed below, please contact the Digital Archives at 509.235.7509.

Record Creator: Eastern Regional Archives

Custodial Partner: Eastern Regional Archives

Submitting Partner: Eastern Regional Archives

Record Series: Adams County Birth Records

Start Year: 1900 End Year: 2010 Access Number: DAN: AU52-03A-01

Introduction: This database contains the names of people who were born in Adams County from 1893-1907, 1910-1915, (several delayed birth returns: 1942). Their names were transcribed from the Adams County Birth Returns.

Citation: Adams County Birth Records

☒ Records are orderable. ☒ Data is archival

Previous Next

Archive This! - Washington State Archives, Digital Archives

Where is Your Data?

Please use the form below and show us where your files are located.

Metadata Format: Comma seperated values (CSV)

☒ The metadata references Digital Objects. ☒ First row contains the name of the columns.

Metadata Path: V:\Sample Data\Fat File\Eastern Region Archives\Adams County Birth\Data\Adams Co. Birth RETUR Browse

Digital Object Format: TIFF

Digital Object Folder: V:\Sample Data\Fat File\Eastern Region Archives\Adams County Birth\Images Browse

Digital Object Restrictions: is displayable on the website

Previous Next

Archive This

Archive This! - Washington State Archives, Digital Archives

Match Your Data Fields To Our Data Fields

Please use the form below to map your data field names to the data field names that the Digital Archives use. If you have questions please contact the Digital Archives at 509.235.7509.

Template: Adams County Birth Records [Load] [Save] [Delete]

	Source Column	Destination Column	Display Name	Restriction Code	Data Type	Order
▶	DOCUMENT NO.	Reference Number	DOCUMENT NO.	Is displayable on the website	Text	1
	CHILD'S LAST ...	Child's Last Name	CHILD'S LAST NAME	Is displayable on the website	Text	2
	CHILD'S FIRST ...	Child's First Name	CHILD'S FIRST NAME	Is displayable on the website	Text	3
	BIRTH DATE	Birth Date	BIRTH DATE	Is displayable on the website	Text	4
	SEX	Child's Gender	SEX	Is displayable on the website	Text	5
	FATHER'S LAST...	Father's Last Name	FATHER'S LAST NA...	Is displayable on the website	Text	6
	FATHERS FIRST...	Father's First Name	FATHERS FIRST NA...	Is displayable on the website	Text	7

File Column Name: Reference Number

☐ File column contains the full name of the file, including the extension.

	DOCUMENT	CHILD'S LAST	CHILD'S FIRST	BIRTH DATE	SEX	FATHER'S LAST	FATHERS FIRST	M
▶	eaadbr1	McCann		Dec 19 1894	Female	McCann	Asa	Wr
	eaadbr10	Dowdy		Aug 17 1895	Male	Dowdy	T	Re
	eaadbr100	Eastman		Apr 15 1899	Male	Eastman	W S	Sta
	eaadbr101	Goodwin		Mar 21 1899	Male	Goodwin	R M	La
	eaadbr102	Weber	Fred, Jr	Mar 25 1899	Male	Weber	Fred	Re
	eaadbr103	Oestreich	Reubin	Mar 20 1899	Male	Oestreich	John	Scl

Previous Next

Archive This

Archive This! - Washington State Archives, Digital Archives

Transfer Data to the Digital Archives

You may transfer the data you have prepared to the Digital Archives by uploading it via SFTP, or by exporting it to a hard drive. Use the form below to send the data to us!

☐ Select All Transfer Items

☒ Adams County Birth Records - 567 record(s) - 758,629,100 bytes

Upload To Digital Archives

Export To Hard Drive

Remove From Queue

Queue More

The Evolution

- Originally WADA Data Transfer
 - Had following Issues
 - No data validation
 - No update capability
 - No upload/export support
 - Column mapping was hard

How is it implemented now?

- Written in C#
- Uses WPF for user interface
- Uses WCF for interaction with the DA database
- File transfer takes place using FTPS (block mode)

Highly Desirables

- Automate application on client's server
- Add a new record series without a deployment and code change

Auto Todd

- Monitors incoming data
- Virus scans new incoming data
- Verifies that new incoming data is the same on our server as it was on the client's server
- Processes new incoming data
- Sends alerts based on process state
- Backs up data to remote storage area
- Tracks all incoming data and all of its states as it travels through the ingestion system

The Evolution

- First iteration used SSIS
- Everything was manual
- Data validation and file matching was done on the Digital Archives' end
- Required much human interaction

How is it implemented now?

- Consists of several Windows services written with custom C# code
- Each service hosts a workflow which outlines how the specific service will work

Main Processing

- Ingests individual records in the archives
- Distributed on sixteen quad core servers and two eight processor servers
- Converts files to presentation format
- Extracts text from audio
- Each record either goes in completely or not at all (atomic)

Contacts

Dan Waterbly

dan.waterbly@sos.wa.gov